

The Evolution of the Major Hepatitis C Genotypes Correlates with Clinical Response to Interferon Therapy

Phillip S. Pang¹*, Paul J. Planet², Jeffrey S. Glenn³*

1 Department of Medicine, Division of Infectious Diseases and Geographic Medicine and Division of Gastroenterology and Hepatology, Stanford University School of Medicine, Palo Alto, California, United States of America, **2** Department of Pediatrics, Division of Infectious Diseases, Columbia Presbyterian Medical Center and, Sackler Institute of Comparative Genomics, American Museum of Natural History, New York, New York, United States of America, **3** Department of Medicine, Division of Gastroenterology and Hepatology, Stanford University School of Medicine and the Palo Alto Veterans Administration Medical Center, Palo Alto, California, United States of America

Abstract

Background: Patients chronically infected with hepatitis C virus (HCV) require significantly different durations of therapy and achieve substantially different sustained virologic response rates to interferon-based therapies, depending on the HCV genotype with which they are infected. There currently exists no systematic framework that explains these genotype-specific response rates. Since humans are the only known natural hosts for HCV—a virus that is at least hundreds of years old—one possibility is that over the time frame of this relationship, HCV accumulated adaptive mutations that confer increasing resistance to the human immune system. Given that interferon therapy functions by triggering an immune response, we hypothesized that clinical response rates are a reflection of viral evolutionary adaptations to the immune system.

Methods and Findings: We have performed the first phylogenetic analysis to include all available full-length HCV genomic sequences ($n = 345$). This resulted in a new cladogram of HCV. This tree establishes for the first time the relative evolutionary ages of the major HCV genotypes. The outcome data from prospective clinical trials that studied interferon and ribavirin therapy was then mapped onto this new tree. This mapping revealed a correlation between genotype-specific responses to therapy and respective genotype age. This correlation allows us to predict that genotypes 5 and 6, for which there currently are no published prospective trials, will likely have intermediate response rates, similar to genotype 3. Ancestral protein sequence reconstruction was also performed, which identified the HCV proteins E2 and NS5A as potential determinants of genotype-specific clinical outcome. Biochemical studies have independently identified these same two proteins as having genotype-specific abilities to inhibit the innate immune factor double-stranded RNA-dependent protein kinase (PKR).

Conclusion: An evolutionary analysis of all available HCV genomes supports the hypothesis that immune selection was a significant driving force in the divergence of the major HCV genotypes and that viral factors that acquired the ability to inhibit the immune response may play a role in determining genotype-specific response rates to interferon therapy.

Citation: Pang PS, Planet PJ, Glenn JS (2009) The Evolution of the Major Hepatitis C Genotypes Correlates with Clinical Response to Interferon Therapy. PLoS ONE 4(8): e6579. doi:10.1371/journal.pone.0006579

Editor: Sheila Mary Bowyer, National Institute for Communicable Diseases, South Africa

Received: February 18, 2009; **Accepted:** July 9, 2009; **Published:** August 11, 2009

Copyright: © 2009 Pang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. JSG funded by a Burroughs Wellcome Fund Clinical Scientist Award in Translational Research, NIH RO1 DK064223, and the Center for Translational Research in Chronic Viral Infections. PSP supported by a Stanford Dean's Fellowship, NIH T32 AI070502, and NRSA 1F32AI082930.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: jeffrey.glenn@stanford.edu

These authors contributed equally to this work.

Introduction

Nearly 170 million people worldwide are chronically infected with hepatitis C virus (HCV) [1]. In the US, HCV is the leading cause of hepatocellular carcinoma and the leading indication for liver transplantation [2]. The standard of care for the treatment of chronic hepatitis C is combination therapy with pegylated interferon and ribavirin. Pegylated interferon (PEG-IFN) is a synthetic variant of interferon- α , a naturally occurring cytokine whose endogenous role is to activate the innate immune response. Injected PEG-IFN is hypothesized to function by mimicking this natural cytokine. Ribavirin (RBV) is a nucleoside analog. It is thought to act through a combination of modalities (as reviewed in [3,4]).

Large clinical trials of PEG-IFN/RBV therapy have revealed significantly different response rates for the various HCV genotypes. There are six major HCV genotypes, numbered 1 to 6. Genotype 2 is the most responsive, with a sustained virologic response (SVR) rate of greater than 80%. Studies also suggest that it is reasonable to treat some patients infected with this genotype for only 12–16 weeks [5,6,7]. Conversely, the most prevalent genotype worldwide, genotype 1, is the least responsive. The SVR rate for patients infected with genotype 1 is less than 50%. Current guidelines recommend 48 weeks of therapy for this genotype; shorter courses of therapy have been demonstrated to be sub-optimal [8].

There currently exists no systematic explanation for these genotype-specific differences in clinical outcome [4,9,10]. It is

assumed that genotype-specific clinical response rates are the result of a confluence of host and viral factors. What specific host factors, human demographics/geographic patterns, and/or viral factors determine interferon response rates remains a challenging area of inquiry. Furthermore, whether factors that govern outcome for one genotype play a similar role in other genotypes remains to be more fully explored.

Numerous laboratory studies suggest that certain viral factors are able to inhibit aspects of the innate immune response (as reviewed in [11,12,13]). These cell culture studies, however, highlight the gap that currently exists between laboratory models and the human host. For example, the HCV replicon system allows for the study of HCV RNA replication in cell culture. Using this system, it was observed that genotype 2 replicons were more resistant to interferon than genotype 1 replicons, the opposite of what is observed clinically [14]. Thus, this observation may be a culture system artifact [15] that highlights the challenge of ascertaining the clinical relevance of findings first discovered in laboratory models.

Humans are the only known natural hosts for HCV, a virus that is estimated to be hundreds and possibly thousands of years old [16,17,18]. This lengthy relationship may have allowed HCV to accumulate adaptive mutations that confer increasing resistance to the human immune system. Interferon therapy functions by activating the innate immune response, which is comprised of direct intracellular defenses such as the PKR, Mx and RNaseL proteins, and innate immune cells, including NK, dendritic, monocyte, macrophage, and NKT cells. Once activated, the innate immune system also plays a critical role in the proper stimulation and coordination of the adaptive immune response [13,19].

We therefore hypothesize that genotype-specific clinical response rates to interferon-based therapies are a reflection of HCV evolutionary adaptations to the immune system. We do not hypothesize that modern interferon therapy itself selected for the various HCV genotypes. Instead, we are hypothesizing that the immune system that is activated by interferon therapy has co-evolved with HCV.

One evolutionary pattern that would strongly indicate that a selective pressure was favoring adaptations to the immune system would be a strict correlation of increasing non-response to treatment with the relative ages of the genotypes—such that, as each new genotype emerged it would have a more resistant phenotype than its ancestor.

HCV was first divided into genotypes by the seminal work of Simmonds and others in 1993 [20,21], based on an analysis of one segment of the HCV genome from 76 different patients (Figure 1A). Evolutionary analysis limited to only portions of a genome, however, can be misleading [18,21,22]. For instance, by analyzing 27 full-length HCV genomes Salemi and colleagues [23] (Figure 1B) found a different phylogenetic pattern for the relationships amongst the six HCV genotypes. Also of note is that neither analysis determined the relative evolutionary ages of the various genotypes.

Determining the relative ages of the major HCV genotypes is critical to testing our hypothesis that a correlation exists between genotype age and clinical resistance. Relative ages can be determined through the use of an outgroup, which roots the phylogeny and establishes the direction of time. In the analysis presented here, we used GB Virus B (GBV-B) as an outgroup. This allowed us to root our HCV phylogeny and establish for the first

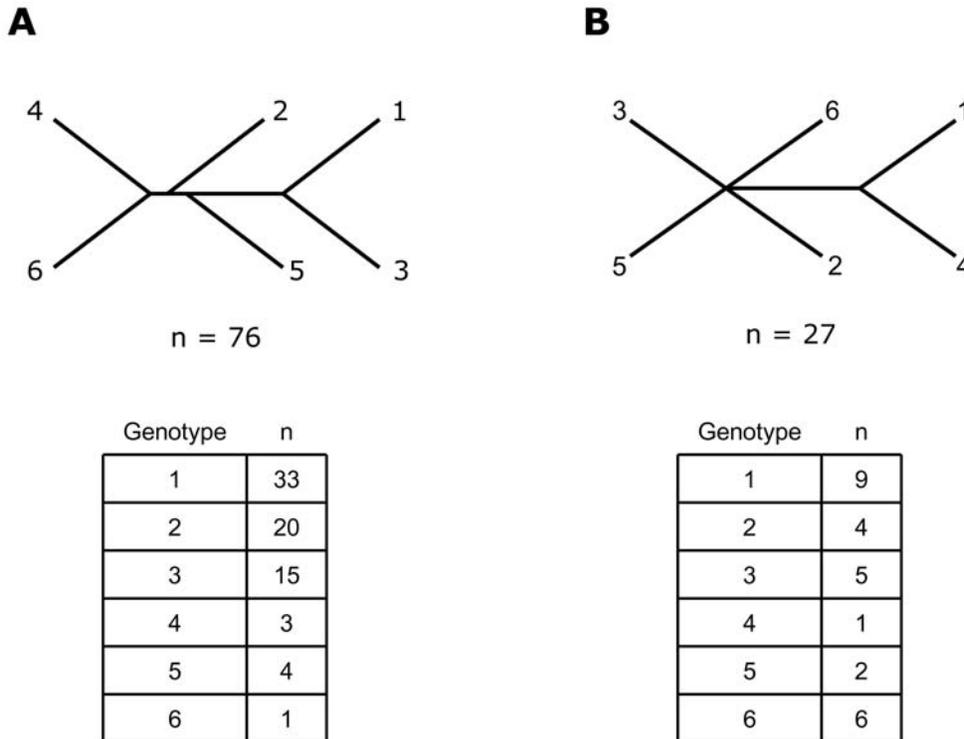


Figure 1. Unrooted HCV Cladograms From Previous Studies. Panel A shows the first cladogram to divide HCV into six genotypes, based on a neighbor-joining analysis of the NS5 region of HCV that included 76 sequences (Simmonds *et al.* 1993). Panel B shows a more recent HCV consensus tree with a different genotype branching pattern compared to Panel A, based on an analysis of 27 full-length HCV genomic sequences (Salemi *et al.* 2002). The table below each panel indicates the genotype distribution of the sequences analyzed in these studies. doi:10.1371/journal.pone.0006579.g001

time the relative ages of the major HCV genotypes. GBV-B was chosen as the outgroup for two reasons: first, GBV-B, a virus that causes hepatitis in New World monkeys, is the closest viral relative of HCV and the only other member of the *hepacivirus* genus. Second, biochemical evidence suggests that proteins in GBV-B share highly-specific functionality with their homologs in HCV [24].

The evolutionary analyses of HCV that have been performed to date have also been based on a limited number of genomic sequences. A prime reason for this is that reliable methods of tree construction, such as maximum likelihood (ML) and maximum parsimony (MP), require considerable amounts of computational power. Thus, often only a subset of available sequences is actually analyzed. For this reason, Salemi *et al.* limited their analysis to 27 genomes.

In this work, we perform a comprehensive analysis of all the >300 genomes found in the European HCV database. This order of magnitude increase in the number of genomes analyzed was made possible by the NSF-funded Cyberinfrastructure for Phylogenetic Research (CIPRES) Project, which allows for web-based access to the San Diego Super Computer facility and newly developed evolutionary algorithms that dramatically reduce computational time.

We thus sought to construct the first evolutionary tree of HCV that incorporated all known genomic sequences and would allow for the relative ages of all HCV genotypes to be determined. We then used this tree to test our hypothesis that clinical resistance to interferon correlates with HCV genotype age. Finally, we used ancestral sequence reconstruction to identify HCV loci that potentially play a role in determining genotype-specific clinical outcomes.

Methods

Sequence Selection and Alignment

All 348 full-length HCV genomic sequences publicly available as of October 2007, were downloaded from the European HCV database [25]. Reflecting its wider prevalence, 236 of these sequences were genotype 1. Forty-five sequences were genotype 2, seven were genotype 3, ten were genotype 4, two were genotype 5, and forty-five were genotype 6. Three sequences were putative recombinants and were discarded.

Two separate methods were used to align the coding regions of these sequences, in order to address potential alignment strategy biases. The first method utilized ClustalX [26] followed by inspection, which ensured that the alignment respected known viral protein properties. The second alignment method, MAFFT [27], was fully automated. Over 90% of the HCV genome encodes a single polyprotein. To ensure in-frame alignment, amino acid sequence alignments were first generated and then used to guide the nucleic acid sequence alignments. For both alignment programs, gap opening and extension parameters were set to their defaults.

GBV-B was selected as the outgroup because it is the closest known viral relative of HCV that is not thought to fall within the HCV clade [24,28,29]. Broadening the outgroup to include GB virus A and GB virus C was not possible because the alignment of these viruses to HCV was poor. Thus, these viral sequences were not used in our analyses.

Phylogenetic Analyses

Our two amino acid alignments and the two nucleotide (NT) alignments, each alignment having been generated by the two alignment methods described above, were then analyzed using ML, MP, and neighbor-joining (NJ) techniques (Figure 2). Additionally, a combined alignment was created by concatenating our amino acid and nucleotide datasets into a single matrix. This combined alignment was then analyzed using the MP method. Combining amino acid sequences with nucleotide sequences has been proposed as a method for extracting as much information from sequences as possible and acting to weigh protein coding amino acid data without *a priori* transformation costs [30].

PAUP [31] was used to perform the NJ analyses. The minimum evolution criterion was used and branch lengths were allowed to be negative except when calculating tree scores. In this situation, branch lengths were set to zero. Ties were broken randomly. The best NJ tree (amino acid dataset) score was -9.09196 ; the best NJ tree (nucleotide dataset) score was -12.83397 .

For MP analyses, the program PAUPRat [32] in conjunction with PAUP [31] was used to perform an aggressive search using the Ratchet method [33]. 200 ratchet iterations were performed for each dataset with random addition (RA) followed by tree branch reconnection (TBR) swapping, randomly upweighting

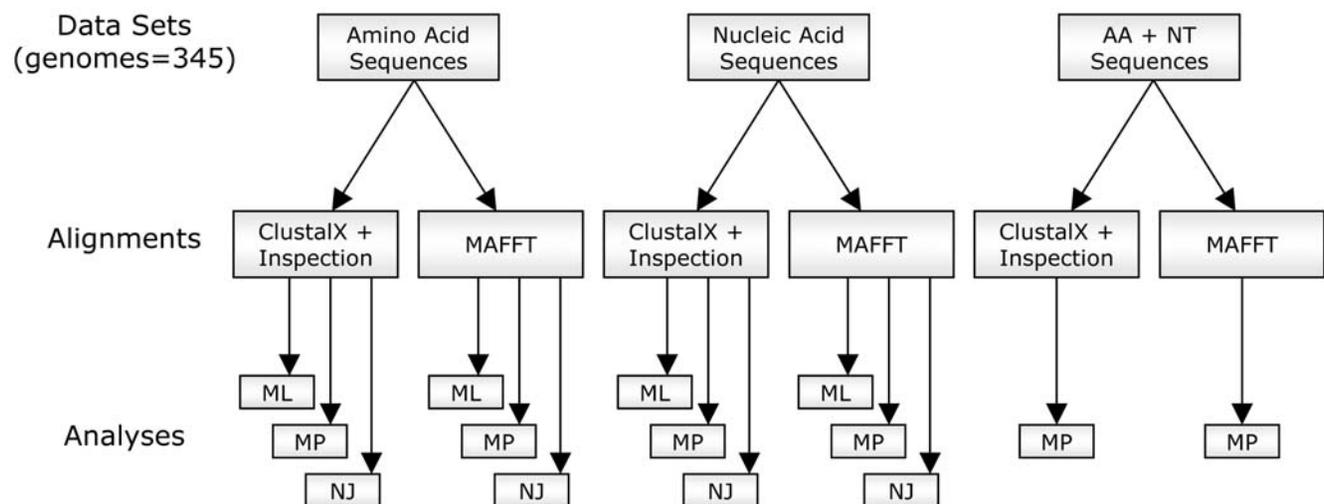


Figure 2. Flow Chart of the Evolutionary Analyses Performed in This Study. ML: maximum likelihood; MP: maximum parsimony; NJ: neighbor joining; MAFFT: multiple sequence alignment based on fast Fourier transform. doi:10.1371/journal.pone.0006579.g002

15% of characters at each iteration and saving only one tree at each iteration. The resulting trees were then used as starting trees for TBR searching using the Multrees option in PAUP. Gaps were treated as a state and all characters and state transformations were weighted equally. Bootstrap values were calculated using 100 bootstrap iterations, using 10 replicates of RA followed by TBR in each iteration [29]. The data set for each iteration was generated by re-sampling (with replacement) the characters in the alignment. Bremer supports [29] were calculated using the program Autodecay [34] in conjunction with PAUP, using 10 TBR replicates for each node in the MP phylogeny. See Table 1 for MP tree scores and statistics.

ML NT analyses were performed using GARLI [35], via the CIPRES portal [36]. Prior ML analyses used a GTR+Gamma site model [37]. We used the program HYPHY [38] to perform both exhaustive and hierarchical model testing based on the likelihood ratio, and determined the optimal model to be GTR+gamma+invariant. Nevertheless, use of invariant sites did not affect the tree structure determined (data not shown). The shape of the gamma function was inferred from the data set. More than 35 separate ML analyses were performed, in order to address stochastic concerns and better ensure that the resultant tree had the optimal maximum likelihood. The optimal NT ML tree -lnL score was 594309. RAxML[39] was used to assess node support with 100 rapid bootstrap replicates. Bootstrap values were drawn on the best-scoring GARLI ML. RAxML was also used for amino acid analyses, using a WAG+Gamma+F model. The optimal AA ML tree -lnL score was 198250. At the time of our analyses, the CIPRES server used RA with no Multrees and no swapping for heuristic tree searching. Out of concern that this less aggressive search strategy might limit our ML analyses, we also used as starting trees the outputs from our more aggressive MP search strategy. Resultant trees were unchanged.

Evaluating Taxon Sampling

All available whole HCV genomes were initially analyzed, based on the assumption that adding taxa increases phylogenetic accuracy [40]. As a consequence, genotype 1 sequences represented 236/345 of the analyzed genomes. We utilized two approaches to address the potential bias introduced by this uneven taxonomic sampling. First, we performed our analyses using three different optimality criteria, since uneven taxon sampling might be expected to cause the various optimality criteria to produce different phylogenetic trees [41,42]. Second, we utilized a taxon jackknifing technique. In our jackknife analyses, genotype 5 sequences were excluded, as only two existed. Each jackknifed dataset was constructed by randomly selecting seven taxa per genotype, without replacement, from the initial alignments. The number seven was selected as this represented the number of taxa in genotype 3, the second least represented genotype. Random selection was performed 10 times, for each of the amino acid,

nucleotide and concatenated alignments, resulting in 30 datasets, each of which consisted of 35 taxa plus the outgroup GBV-B. All of the above described analyses were then repeated on these datasets.

Clinical Trial Data Compilation

A PubMed search for English language, prospective trials that studied the effect of PEG-IFN/RBV combination therapy for the treatment of chronic hepatitis C was used to identify relevant trials. Published reviews were also consulted to ensure comprehensiveness [8,43]. Studies published (including early e-publication) up to December 2007 were included. The fact that sustained virologic response rates vary according to genotype has already been well-established [5,8,43,44,45,46]. For illustrative purposes, were therefore compiled here only the major, large prospective trials. Specifically, early trials with less than 100 patients in total or arms with less than 25 patients were excluded, except in the case of genotype 4 for which there primarily exist a limited number of smaller trials. For genotype 4, trials with less than 25 patients were excluded. Published non-inferiority trials comparing different PEG-IFN formulations were also excluded, as they assessed only end-of-treatment virologic response. The genotype-specific clinical response rates shown are based on intention-to-treat analyses, and represent averages weighted according to the number of patients in each indicated study (Table 2). Notably, clinically relevant differences in outcome among HCV sub-genotypes (subtypes) [8] have not been observed. Our approach was therefore limited to genotype-level outcome data.

Testing Correlation

Standard tests of correlation are not appropriate for items that are related by descent in an evolutionary hierarchy [47,48]. We therefore used Shimodaira's approximately unbiased (AU) tree-based statistical test [49,50] to evaluate the critical nodes within our tree topology and test the statistical significance of our observed correlation between relative genotype age and SVR. Specifically, we constrained the clade including genotypes 1 and 4 not to exist in our optimal tree ($T_{no1,4}$). We then constrained the clade including genotypes 1, 3, 4, and 5 to not exist in our optimal tree ($T_{no1,4,3,5}$). Each of these alternate trees, which disrupt the evolutionary branching pattern of the major genotypes, was then tested against the optimal ML tree using Shimodaira's AU test with the RELI bootstrap approximation (1000 replicates), to determine if they were statistically inferior. These analyses were carried out using the program CONSEL [51].

Identification of Viral Resistance Loci

Fitch optimization [52] was used to reconstruct the ancestral protein sequence at each of the three nodes that lie on the main trunk of our tree [53,54]. These ancestral sequences were then screened, in two steps, to identify the positions that correlate with

Table 1. Parsimony Tree Statistics.

	Optimal Score (Clustal)	No. of Trees Found (Clustal)	Consistency Index	Retention Index	Rescaled Consistency Index	Consensus Fork Index	Rohlf Consensus Index	Total No. of Characters (Parsimony Informative)
AA+NT	175615	228	0.162	0.76	0.123	0.901	0.806	12881 (8107)
NT	140497	32	0.125	0.752	0.094	0.977	0.959	9669 (6227)
AA	3334	31104	0.293	0.799	0.234	0.936	0.94	3212 (1880)

doi:10.1371/journal.pone.0006579.t001

Table 2. Prospective HCV Trials of Therapy with Pegylated Interferon and Ribavirin.

Genotype	Avg. SVR	Therapy Duration	Patients	Study Specific SVR ^a	
1	46%	48 wks	917	Hadziyannis <i>et al.</i> 2004 [44]	52% (141/271)
				Fried <i>et al.</i> 2002 [102]	46% (137/298)
				Manns <i>et al.</i> 2001 [45]	42% (145/348)
2	82%	12–24 wks	788	Dalgard <i>et al.</i> 2008 ^b [103]	97% (30/31)
				Shiffman <i>et al.</i> 2007 ^b [6]	75% (268/356)
				Yu <i>et al.</i> 2007 [7]	95% (142/150)
				Mangia <i>et al.</i> 2005 [5]	80% (171/213)
				Von Wagner <i>et al.</i> 2005 [104]	92% (35/38)
3	72%	12–24 wks	657	Dalgard <i>et al.</i> 2008 ^b [103]	92% (106/115)
				Shiffman <i>et al.</i> 2007 ^b [6]	66% (244/369)
				Mangia <i>et al.</i> 2005 [5]	66% (46/70)
				Von Wagner <i>et al.</i> 2005 [104]	73% (75/103)
4	60%	24–48 wks	676	Kamal <i>et al.</i> 2007 [46]	63% (239/378)
				Derbala <i>et al.</i> 2005 [105]	29% (10/35)
				Kamal <i>et al.</i> 2005 [106]	70% (48/69)
				El-Zayadi <i>et al.</i> 2005 [107]	55% (22/40)
				Alfaleh <i>et al.</i> 2004 [108]	43% (12/28)
				Hasan <i>et al.</i> 2004 [109]	68% (45/66)
				Shobokshi <i>et al.</i> 2003 ^c	50% (30/60)
5	No prospective trials of pegylated interferon and ribavirin				
6	No prospective trials of pegylated interferon and ribavirin				

^aSustained Virologic Response: Number of patients with no detectable virus 6 months after completion of therapy divided by the total number of patients treated, based on intention-to-treat analyses.

^bData from the 24 week arm of the study.

^c2003 AASLD Abstract # 996.

doi:10.1371/journal.pone.0006579.t002

increased resistance. First, because resistance adaptations result from sequence mutation, these ancestral sequences were screened to identify residues that had undergone mutation as HCV evolved along the main trunk of the tree. These positions were then screened to identify positions that, having mutated, conserved the particular mutation in all future progeny. This second screening step is based on the assumption that functionally advantageous mutations are likely to be conserved.

Results

The results of each individual phylogenetic analysis are shown in Figure 3. Regardless of whether amino acid or nucleotide sequence data was analyzed, or which of the two alignment methods were utilized, striking concordance was found to exist among the three phylogenetic inference methods. Individual optimal trees from each analysis were also supported by bootstrap values and/or Bremer decay indices, and differed only in the placement of genotypes 5 and 6. In all but one case (parsimony nucleotide analysis) the relative branching order of those genotypes with prospective clinical outcome data (i.e. genotypes 1, 2, 3, and 4) was identical.

The end result of our phylogenetic analyses is summarized in figure 4, left panel. It reveals that HCV genotype 2 branched first, genotypes 1 and 4 branched last, and genotypes 3, 5, and 6 branched sometime after genotype 2 but before genotype 4. Using population genetic methods to analyze limited portions of the genome from a sampling of genotypes (1, 4, and 6), Pybus and colleagues [16,18] estimated origin times for genotypes 1, 4 and 6 that are notably

concordant with the more complete branching order we have determined. Other studies [17,18,55] have inferred the absolute age of certain HCV sub-genotypes/subtypes (e.g. 1a, 1b and 3a) and found them to be relatively recent; these findings do not contradict the relative ages we have inferred for the major genotypes.

To gauge the effect of taxon sampling on our phylogenetic analysis, we performed the taxon jackknife technique described above, in which we randomly selected an equal number of taxa from each genotype for repeated analyses. All jackknife replicates using the amino acid data, for all optimality criteria (ML, NJ, and MP), gave exactly the same genotype branching order seen in Figure 4. For nucleotide data, jackknife replicates analyzed using NJ and MP methods also resulted in the same branching order. Likewise, the combined amino acid and nucleotide datasets analyzed using MP criteria were 100% concordant with the overall consensus. The ML analyses of the jackknifed nucleotide data sets resulted in inconsistent branching patterns. Only one ML NT jackknifed data set resulted in a tree with the same branching pattern as seen in our overall consensus. The remaining nine differed amongst each other in the location of the root and/or overall topology, with most rooting within or near genotype 6.

A compilation of the outcome data from 19 prospective trials of combination therapy with PEG-IFN/RBV to treat chronic hepatitis C is shown in Table 2. As has been previously observed [5,8,43,44,45,46], genotype-specific response rates are hierarchical. From highest to lowest, the pattern of SVR rates is: genotype 2 > genotype 3 > genotype 4 > genotype 1. There currently exist no prospective trials of PEG-IFN/RBV therapy for genotypes 5 and

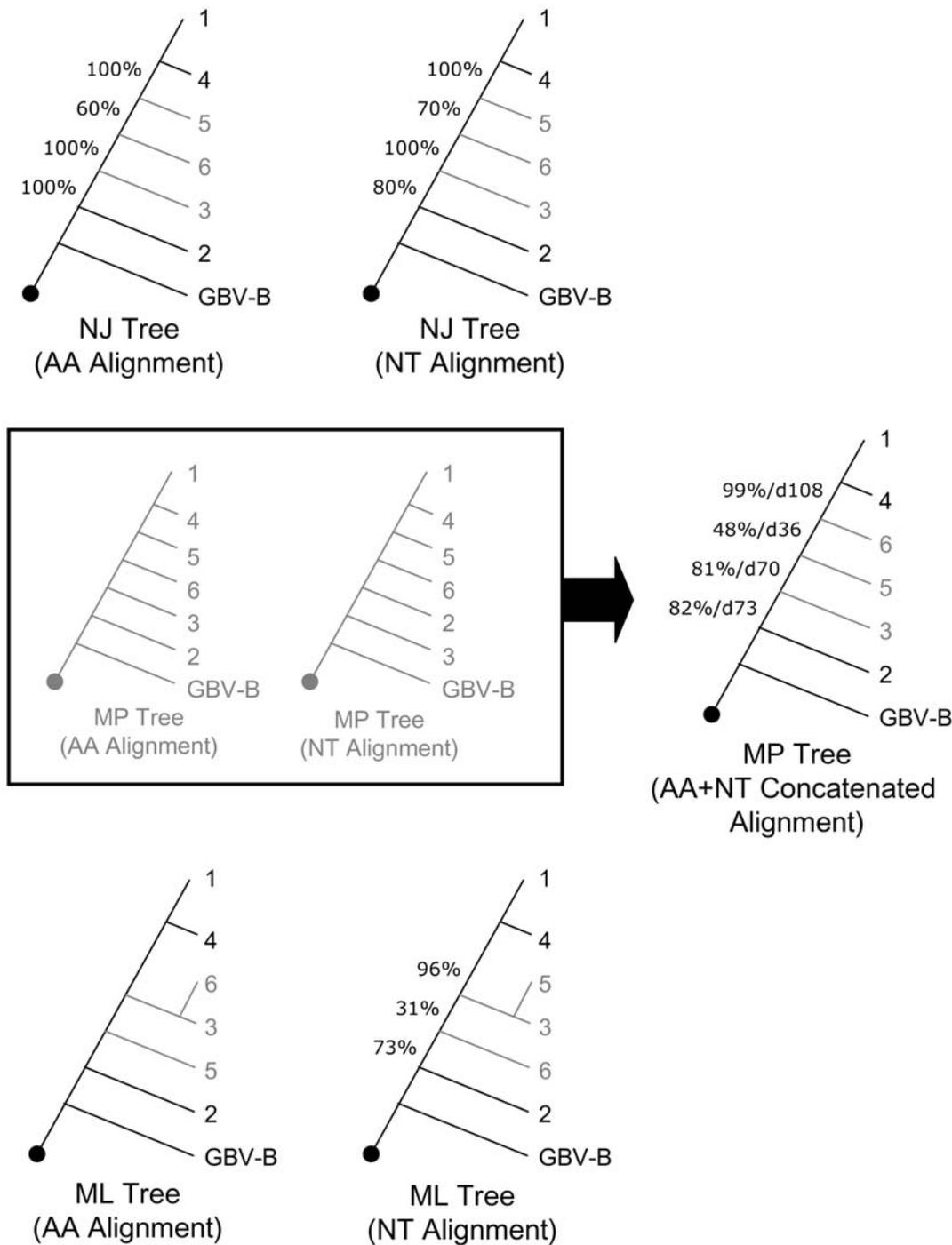


Figure 3. Rooted Neighbor-Joining (NJ), Maximum Parsimony (MP), and Maximum Likelihood (ML) cladograms depicting the evolution of the major hepatitis C virus genotypes. A nexus file of complete tree data is available online (Dataset S1). Numbers represent bootstrap support and Bremer decay indices. AA: amino acid sequences; NT: nucleotide sequences. doi:10.1371/journal.pone.0006579.g003

6. Notably, no compelling evidence exists that suggests that sub-genotypes (i.e. 1a versus 1b) have clinically relevant differences in outcome [8], though evidence exists that sequence variations within subtypes may affect clinical outcome [56,57]. Therefore, the present study was restricted to a genotype-level analysis of clinical response and sequence evolution.

When genotype-specific clinical response rates were mapped onto our phylogenetic tree, a correlation between genotype age

and clinical resistance was revealed (Figure 4). As hypothesized, early branching genotypes were noted to have the best clinical outcomes and require the least duration of therapy, while genotypes that branched later (i.e. more recently) have higher rates of clinical resistance and need to be treated for much longer. Therefore, each newly emerged genotype has greater resistance than its ancestor, indicating the likely presence of a selective pressure favoring resistance. To test the statistical significance of

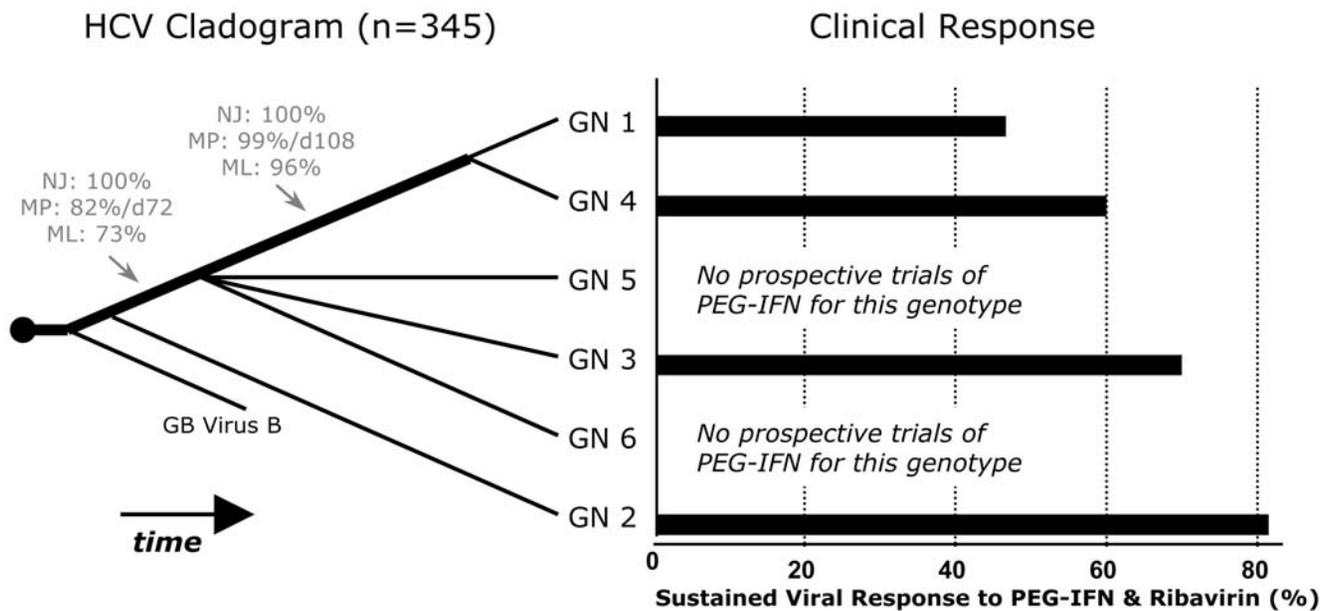


Figure 4. Rooted Consensus Cladogram Resulting From An Analysis of 345 Full-Length HCV Genomic Sequences. The evolution of the major (HCV) genotypes and their correlation to clinical outcome is depicted. Values denote bootstrap support and Bremer decay indices for the indicated phylogenetic inference method. doi:10.1371/journal.pone.0006579.g004

this correlation between relative genotype age and SVR, we applied Shimodaira's AU test (see Methods). Briefly, we searched for the best possible trees that disrupted the observed branching pattern and compared their likelihoods to the likelihood of our optimal tree. In every case, trees that contradicted our tree topology and disrupted the observed correlation with interferon susceptibility were significantly suboptimal ($P < 0.01$). Thus, by this rigorous standard, the observed correlation is statistically significant. A far more conservative estimate of the significance of this correlation was calculated using the equation, $N = (2n-3)! / (2n-2(n-2)!)$, where n is the number of taxa (in this case, genotypes are the taxa), and N equals the number of ways in which taxa can be ordered on a branching tree. Thus there were a total of $N = 15$ possible ways in which the 4 genotypes with known clinical outcomes could be ordered on our tree, making the likelihood of a tree pattern that matched SVR outcomes, purely by chance, equal to 1 out of 15.

This correlation between clinical resistance and branching order allows for the following prediction: prospective clinical trials of genotypes 5 and 6 using PEG-IFN/RBV, for which prospective data is currently lacking, will likely show intermediate SVR rates, similar to genotype 3. Antaki and colleagues have published a retrospective study of genotype 5 infected patients treated with PEG-IFN/RBV. By this retrospective analysis, genotype 5 infected patients have an SVR rate of 67% [58], exactly as our tree predicts.

Ancestral sequence reconstruction, as described in Methods, was then used to identify viral elements potentially responsible for this clinical resistance trend. Our analysis resulted in 55 hot spots: positions that mutated as HCV became increasingly resistant. The locations of these hotspots are shown in Figure 5. N1 denotes positions associated, by reconstruction analysis, with genotype 1 being more resistant than any other genotype; N2 denotes positions associated with genotypes 1 and 4 being more resistant than genotypes 2, 3, 5, and 6; N3 denotes positions associated with genotypes 1, 3, 4, 5, and 6 being more resistant than genotype 2.

These hotspots were then binned in groups of 50 amino acids, resulting in the frequency histogram shown at the top of Figure 5. This histogram identifies the two 50 amino acid regions of the HCV genome that contain the greatest number of hotspots. These two regions fall within the PKR binding domains of the E2 and NS5A proteins of HCV [59,60], suggesting their potential importance as viral factors that may determine genotype-specific responses to interferon therapy.

Discussion

In this study, the relative evolutionary age of HCV genotypes has been shown to correlate with the likelihood of a sustained virologic response to interferon-based therapies (Figure 4). We suggest that the observed correlation stems from consistent selective pressure generated by the host immune system. However, a number of alternative explanations could account for the observed correlation between response to interferon-based therapies and relative genotype age. It is possible that the observed correlation is a product of chance, bias, or analytical limitations. Alternatively, the branching pattern may be accurate, but the patterns may not be directly causally related. We discuss each of these possibilities below, as well as the current experimental evidence in favor of immune-mediated selection.

With regard to chance, bias, or potential analytical limitations, our results are dependent on the reliability of evolutionary reconstruction techniques and limitations of the data set. We have attempted to address these concerns through the use of three different techniques (NJ, ML, and MP), two alignment methods, and rigorous statistical testing. Our approach included performing bootstrap and Bremer decay index analyses to measure support for branches. Importantly, we note that neither the small discrepancies that exist between phylogenetic techniques, nor areas of weaker support, contradict the overall branching pattern of the tree.

As many available whole HCV genomes as possible were analyzed based on the assumption that adding taxa increases

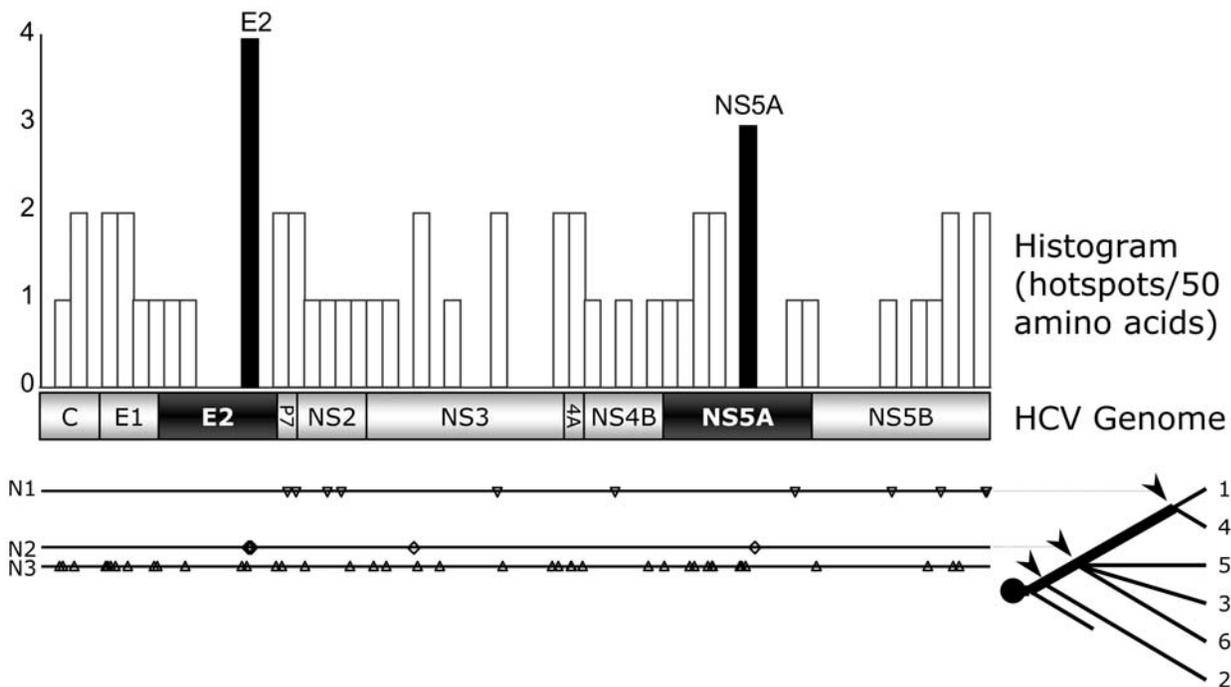


Figure 5. HCV Resistance Loci. Positions (“hotspots”) in the HCV genome that appear to have undergone directed change with respect to immune resistance. N1: positions associated, using ancestral sequence reconstruction techniques, with genotype (GN) 1 being the most resistant. N2: positions associated, using ancestral sequence reconstruction techniques, with GNs 1 and 4 being more resistant than GNs 2, 3, 5, and 6. N3: positions associated, using ancestral sequence reconstruction techniques, with GNs 1, 3, 4, 5, and 6 being more resistant than GN 2. A hotspot histogram (binned in groups of 50 amino acids) is shown at the top. The 10 proteins encoded by the HCV genome are also illustrated. Two loci were identified with the highest density of hotspots (black bars); these loci map to the PePHD domain of E2 and the PKR binding domain of NS5A, both of which have been shown to inhibit the innate immune factor PKR.
doi:10.1371/journal.pone.0006579.g005

phylogenetic accuracy [40]. However, since genotype 1 has been more well-sampled than other genotypes, this assumption may have introduced a bias due to uneven taxonomic sampling. Such uneven taxon sampling might be expected to cause the various optimality criteria used here to produce different phylogenetic trees [41,42]; however, we have shown that they essentially do not. One could further argue that all phylogenetic optimality criteria may have been affected by uneven taxon sampling bias in the same way, leading each to converge on the wrong answer. To address this possibility, we utilized a taxon jackknifing technique, in which an equal number of taxa were randomly selected from each of the genotypes for analysis. This random selection process was repeated 10 times for each type of dataset. For all dataset types, for all inference methods (except the ML analyses of the NT datasets), and for all jackknife replicates, exactly the same genotype branching pattern was obtained. Thus, the observed genotype branching pattern appears robust both to reducing the number of taxa and to equalizing the number of genotype representatives.

Sparse taxon sampling, as opposed to uneven taxon sampling, may have resulted in greater uncertainty of branch placement for less represented genotypes, such as genotype 5, for which only 2 full-length sequences were available. This uncertainty (as reflected in figure 3), led to the inability to determine a definitive relationship of genotype 5 to genotypes 3 and 6. We could only determine that it branched after genotype 2, and before genotypes 1 and 4.

The above tests control for the possibility that the branching pattern of the tree itself was the result of analytical limitations, bias, or chance. To gauge the validity of the observation that this branching order correlates with clinical response to interferon, we

used a battery of comparisons based on Shimodaira’s AU test. Notably, the AU test showed that alternative tree topologies that disrupt the observed correlation are statistically highly unlikely.

With regard to whether this correlation is the result of a causal relationship, we note that our data satisfies a number of criteria commonly used to address the plausibility of causality in statistical analyses (i.e. the Hill criteria [61]). First, we show a biological gradient of responsiveness that correlates with increasing genotype age, with all four genotypes that have well-defined clinical outcomes adhering to this pattern (Figure 4). Second, ancestral sequence reconstruction was used to identify viral proteins putatively involved in genotype-specific immune resistance. These same proteins have been independently identified in biochemical analyses [36,37] to have the ability to inhibit the innate immune response. Thus, the observed correlation has mechanistic plausibility and is consistent with other evidence. Third, biochemical experimentation has been able to reproduce some of the biological gradient we predict. The ability of the E2 and NS5A proteins to inhibit PKR is genotype-specific and consistent with our evolutionary tree [59,60]. An analysis of host cleavage of viral RNA by the RNaseL defense pathway is also consistent with our observed trend [62,63]. For these reasons, it is reasonable to hypothesize that a causal relationship explains our observations, while recognizing that further experimental evidence is needed.

Possible Mechanisms of Immune Selection

Our phylogenetic results are consistent with our hypothesis that a selective pressure generated by the host immune system has played a significant role in HCV evolution and the divergence of genotypes. The specific combination of interferon-induced

immune pathways responsible for this evolution, however, remains largely a matter of speculation.

For example, given that interferon can cause an up-regulated, multi-specific, HCV-specific CD4+ T-cell response [64], one possibility is that genotype-specific differences in clinical outcome are the result of differences in antigenicity [65]. Patients who progress to chronic infection compared to those that resolve their acute infection have increased viral diversity in their envelope gene E2. Thus, diversity may be a product of the emergence of antibody-selected escape variants [66]. HCV escape variants selected by CD8+ T cells, which recognize other regions of the virus, have also been observed [67,68,69,70]. It remains to be determined, however, whether these observed micro-evolutionary processes are the same macro-processes that produced genotype divergence.

An alternative and non-mutually exclusive possibility is that HCV genotype-specific differences in clinical outcomes are the result of the ability to cause immune dysfunction. Patients with up-regulated interferon stimulated genes (ISG) prior to therapy are likely to be non-rapid responders [71,72]. One possible interpretation of this data is that HCV causes immune dysfunction. ISGs are up-regulated but non-functioning or uncoordinated in their response [12]. Notably, an up-regulated ISG state was found to be more prevalent in more refractory genotypes, such as genotype 1 and 4, than 2 and 3 [72].

Yet another possibility is that genotype-specific outcomes are the result of direct inhibition of the intracellular immune pathways within liver cells. One key interferon intracellular pathway activates double-stranded RNA-dependent protein kinase (PKR). PKR shuts down protein production in infected cells, preventing them from being used as factories for virus production. Biochemical and cell culture studies have determined that HCV proteins E2 and NS5A are capable of inhibiting PKR activation [59,60]. Notably, the motif of E2 that we identified as having key ancestral sequence changes that correlated with genotype immune differentiation is the same region identified by biochemical analysis to be critical for E2 inhibition of PKR—the PKR-eIF2 α phosphorylation homology domain (PePHD). Critically, the ability of E2 to inhibit PKR was found to be specific to genotype 1 [59], the genotype most strongly associated with non-response to therapy [8]. Similarly, our ancestral sequence reconstruction analysis identified the same NS5A locus determined by biochemical studies to be important for NS5A inhibition of PKR (PKR binding domain) [60]. Shimotohno and colleagues have observed that the ability of NS5A to inhibit PKR is genotype specific: NS5A from genotype 1 is able to inhibit the innate immune factor PKR, but the NS5A protein from genotype 2 cannot [73]. This biochemical observation can now be explained by our evolutionary analysis.

Another defense mechanism activated by interferon is the RNaseL pathway. When activated, RNaseL cleaves viral RNA. Evidence suggests that genotype 1 is the most resistant of all genotypes to RNaseL, while genotypes 2 and 3 are these least resistant [62,63]. Such differences in RNaseL susceptibility are likely mediated by differences in nucleotide composition. This data is also consistent with our observed correlation between relative genotype age and clinical responsiveness. Notably, a number of other viral factors in HCV that help it overcome the immune response (as reviewed in [12,13]), have been described. The vast majority of these studies, however, involved only HCV genotype 1. Extending such studies to other genotypes would shed much light on the immunobiology of this virus. Such studies may also eventually reveal that PKR inhibition and RNaseL evasion are only a subset of the genotype-specific defense mechanisms that have evolved as a result of immune selection.

Origins of Current HCV Diversity

Our results raise an important secondary question: Why have the more interferon sensitive genotypes (e.g. genotype 2) not become equally resistant over time, especially given the high mutation rate of HCV? At least four, non-mutually exclusive possibilities might explain a pattern in which certain descendants have a beneficial phenotype while other relatives do not.

First, the beneficial trait could have evolved in a specific historical epidemiological context. Given the phylodynamic pattern of HCV [74], this is quite possible. The opportunity for early HCV genotypes to acquire certain traits may have already passed.

Second, the beneficial trait could be relatively difficult to acquire from a genetic perspective. Although HCV has a high mutation rate, a high mutation rate does not necessarily imply that certain traits will be easily acquired (as reviewed in detail by Smith and Simmonds [75]). For HCV specifically, immune resistance may be difficult to acquire due to structure-functional restrictions on the mutagenic space that current HCV strains can explore. Studies have shown that all of the HCV non-structural proteins interact with one another [76]. Mutation is thus limited to non-critical regions or would require multiple, simultaneous compensatory mutations that preserve function. Mutation space in HCV is further constrained by genome ordered RNA structure (GORS) [77]. Significant RNA structure has been found to underlie the entire HCV polyprotein, which would further limit the ability of HCV to mutate without disrupting critical structure-function relationships. [65,75].

The existence of structure-function limitations on the genetic diversity of HCV is clear from the fact that mutations in HCV are not homogeneously distributed [65,69,75,78,79,80]. Mutations are over-represented in defined hypervariable regions such as the N-terminal half of E2, which is where antibodies have been found to bind [66,81]. Conversely, the C-terminal region of E2 that we and others identify as significant for genotype-specific inhibition of PKR is quite static [82,83,84].

The available clinical data also provides indirect support for the possibility that resistance to interferon is not easily acquired over the time scale of treatment. To date, the most likely causes for therapeutic failure are non-compliance, insufficient drug levels or dose reductions, drug toxicity, and interruptions in therapy—not resistant, escape variants [85,86]. In fact, a study of the sequence variability in the NS5A gene during treatment found no selection of interferon resistant HCV strains [87]. Another study of the full length polyprotein also found no significant difference in the number of mutations between non-responders and relapsers [88]. Although accelerated mutational change with interferon treatment was observed to occur in another study, such changes were notably not related to treatment duration and therefore were not felt to explain treatment non-response [89]. The available clinical evidence, therefore, does not contradict the possibility that resistance to interferon may be genetically difficult to acquire.

A third possible reason why different susceptibilities still exist is that resistance to the immune system may come at some cost or interfere with some other selective advantage. It is possible that the earlier branching genotypes possess specialized functions that we do not yet fully recognize, which compensate for their lack of ability to interfere with the immune system. This possibility is supported by reports of clinical differences in the course of disease for different genotypes [90,91,92].

Fourth, it is possible that early branching genotypes maintain higher susceptibilities because of host immune heterogeneity—i.e. through diversifying selection. Numerous host factors have been shown to effect clinical outcome [93]. One example is the

observation that specific host NK-KIR receptor-HLA combinations have been shown to be a determinant of whether acute HCV infection resolves or progresses to chronic infection [94]. A second example of host heterogeneity is the observation that the binding of interferon to its cellular receptor triggers a far more attenuated transcriptional response in African-Americans compared to Caucasians [95]. The SVR pattern of HCV may be a product of all of these phenomena.

Complexities of HCV Biology and Evolution

We suggest that the human immune system has played a significant role in shaping the evolution of HCV. However, it seems highly unlikely that the evolution of HCV has been driven by a single factor. In fact, other studies have emphasized demographic and spatial history as the predominant forces in HCV evolution [74]. This may explain why genotype 6 is endemic to South East Asia while genotype 4 is endemic to central Africa and the Middle East [18,65,96,97,98]. Such phylogeographic explanations are not necessarily inconsistent with our results. Epidemiological processes and immune selection may both have played a role in HCV evolution. For instance, spread of HCV between host populations with inherent immune differences could have been a major evolutionary determinant.

We also acknowledge that this evolutionary analysis does not explain why the treatment of acute HCV is genotype independent [99,100]. We can only speculate that when acutely infected patients are treated, their disease may not yet have established a firm foothold, and that treatment overwhelms any viral inhibition abilities. It may also be the case that chronic infection allows for viral dissemination to more immune-protected regions of the host [101].

References

- (2002) National Institutes of Health Consensus Development Conference Statement: Management of hepatitis C 2002 (June 10–12, 2002). *Gastroenterology* 123: 2082–2099.
- Verna EC, Brown RS, Jr. (2006) Hepatitis C virus and liver transplantation. *Clin Liver Dis* 10: 919–940.
- Hofmann WP, Herrmann E, Sarrazin C, Zeuzem S (2008) Ribavirin mode of action in chronic hepatitis C: from clinical use back to molecular mechanisms. *Liver Int* 28: 1332–1343.
- Feld JJ, Hoofnagle JH (2005) Mechanism of action of interferon and ribavirin in treatment of hepatitis C. *Nature* 436: 967–972.
- Mangia A, Santoro R, Minerva N, Ricci GL, Carretta V, et al. (2005) Peginterferon alfa-2b and ribavirin for 12 vs. 24 weeks in HCV genotype 2 or 3. *N Engl J Med* 352: 2609–2617.
- Shiffman ML, Suter F, Bacon BR, Nelson D, Harley H, et al. (2007) Peginterferon alfa-2a and ribavirin for 16 or 24 weeks in HCV genotype 2 or 3. *N Engl J Med* 357: 124–134.
- Yu ML, Dai CY, Huang JF, Hou NJ, Lee LP, et al. (2007) A randomised study of peginterferon and ribavirin for 16 versus 24 weeks in patients with genotype 2 chronic hepatitis C. *Gut* 56: 553–559.
- Zeuzem S (2004) Heterogeneous virologic response rates to interferon-based therapy in patients with chronic hepatitis C: who responds less well? *Ann Intern Med* 140: 370–381.
- Katze MG, He Y, Gale M, Jr. (2002) Viruses and interferon: a fight for supremacy. *Nat Rev Immunol* 2: 675–687.
- Chisari FV (2005) Unscrambling hepatitis C virus-host interactions. *Nature* 436: 930–932.
- Gale M, Jr., Foy EM (2005) Evasion of intracellular host defence by hepatitis C virus. *Nature* 436: 939–945.
- Sklan EH, Charuorn P, Pang PS, Glenn JS (2009) Mechanisms of HCV survival in the host. *Nat Rev Gastroenterol Hepatol* 6: 217–227.
- Szabo G, Chang S, Dolganiuc A (2007) Altered innate immunity in chronic hepatitis C infection: cause or effect? *Hepatology* 46: 1279–1290.
- Miyamoto M, Kato T, Date T, Mizokami M, Wakita T (2006) Comparison between subgenomic replicons of hepatitis C virus genotypes 2a (JFH-1) and 1b (Con1 NK5.1). *Intervirology* 49: 37–43.
- Bartenschlager R, Pietschmann T (2005) Efficient hepatitis C virus cell culture system: what a difference the host cell makes. *Proc Natl Acad Sci U S A* 102: 9739–9740.
- Pybus OG, Charleston MA, Gupta S, Rambaut A, Holmes EC, et al. (2001) The epidemic behavior of the hepatitis C virus. *Science* 292: 2323–2325.
- Smith DB, Pathirana S, Davidson F, Lawlor E, Power J, et al. (1997) The origin of hepatitis C virus genotypes. *J Gen Virol* 78 (Pt 2): 321–328.
- Pybus OG, Barnes E, Taggart R, Lemey P, Markov PV, et al. (2009) Genetic history of hepatitis C virus in East Asia. *J Virol* 83: 1071–1082.
- Chevaliez S, Pawlotsky JM (2009) Interferons and their use in persistent viral infections. *Handb Exp Pharmacol*. pp 203–241.
- Simmonds P, Holmes EC, Cha TA, Chan SW, McOmish F, et al. (1993) Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. *J Gen Virol* 74 (Pt 11): 2391–2399.
- Simmonds P, Bukh J, Combet C, Deleage G, Enomoto N, et al. (2005) Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology* 42: 962–973.
- Planet PJ (2006) Tree disagreement: measuring and testing incongruence in phylogenies. *J Biomed Inform* 39: 86–102.
- Salemi M, Vandamme AM (2002) Hepatitis C virus evolutionary patterns studied through analysis of full-genome sequences. *J Mol Evol* 54: 62–70.
- Chen Z, Benureau Y, Rijnbrand R, Yi J, Wang T, et al. (2007) GB virus B disrupts RIG-I signaling by NS3/4A-mediated cleavage of the adaptor protein MAVS. *J Virol* 81: 964–976.
- Combet C, Garnier N, Charavay C, Grando D, Crisan D, et al. (2007) euHCVdb: the European hepatitis C virus database. *Nucleic Acids Res* 35: D363–366.
- Thompson JD, Higgins DG, Gibson TJ (1997) CLUSTAL X multiple sequence alignment program. 1.63 ed. Hamburg, Germany: European Molecular Biology Organization.
- Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* 33: 511–518.
- Muerhoff AS, Leary TP, Simons JN, Pilot-Matias TJ, Dawson GJ, et al. (1995) Genomic organization of GB viruses A and B: two new members of the Flaviviridae associated with GB agent hepatitis. *J Virol* 69: 5621–5630.
- DeSalle R, Giribet G, Wheeler W (2002) *Techniques in molecular systematics and evolution*. Basel ; Boston: Birkhäuser. ix, 407pp.
- Agosti D, Jacobs D, DeSalle R (1996) On combining protein sequences and nucleic acid sequences in phylogenetic analysis: the homeobox protein case. *Cladistics* 12: 65–82.
- Swofford DL (1998) PAUP*: Phylogenetic analysis using parsimony (*and other methods). 4.0 beta ed. Sunderland, Massachusetts: Sinauer.
- Sikes DS, Lewis PO (2001) PAUPRat: A tool to implement Parsimony Ratchet searches using PAUP*. Storrs, CT.

33. Nixon KC (1999) The Parsimony Ratchet, a New Method for Rapid Parsimony Analysis. *Cladistics* 15: 407–414.
34. Eriksson T (1997) Autodecay. 2.9.9 ed. Stockholm: Hypercard stack distributed by the author Botaniska institutionen, Stockholm University.
35. Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. University of Texas at Austin.
36. Cyberinfrastructure for Phylogenetic Research (CIPRES) project.
37. Magiorkinis G, Ntziora F, Paraskevis D, Magiorkinis E, Hatzakis A (2007) Analysing the evolutionary history of HCV: puzzle of ancient phylogenetic discordance. *Infect Genet Evol* 7: 354–360.
38. Pond SL, Frost SD, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21: 676–679.
39. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.
40. Zwickl DJ, Hillis DM (2002) Increased Taxon Sampling Greatly Reduces Phylogenetic Error. *Systematic Biology* 51: 588–598.
41. Felsenstein J (1978) Cases in which Parsimony or Compatibility Methods Will be Positively Misleading. *Systematic Zoology* 27: 401–410.
42. Kolaczowski B, Thornton JW (2004) Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. *Nature* 431: 980–984.
43. Kamal SM (2007) Improving Outcome in Patients With Hepatitis C Virus Genotype 4. *Am J Gastroenterol*.
44. Hadziyannis SJ, Sette H Jr, Morgan TR, Balan V, Diago M, et al. (2004) Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. *Ann Intern Med* 140: 346–355.
45. Manns MP, McHutchison JG, Gordon SC, Rustgi VK, Shiffman M, et al. (2001) Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. *Lancet* 358: 958–965.
46. Kamal SM, El Kamary SS, Shardell MD, Hashem M, Ahmed IN, et al. (2007) Pegylated interferon alpha-2b plus ribavirin in patients with genotype 4 chronic hepatitis C: The role of rapid and early virologic response. *Hepatology* 46: 1732–1740.
47. Felsenstein J (1985) *Phylogenies and the Comparative Method*. pp 1.
48. Harvey PH, Pagel MD (1991) *The comparative method in evolutionary biology*. Oxford; New York: Oxford University Press.
49. Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. *Syst Biol* 51: 492–508.
50. Planet PJ (2006) Tree disagreement: Measuring and testing incongruence in phylogenies. *Journal of Biomedical Informatics* 39: 86–102.
51. Shimodaira H, Hasegawa M (2001) CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17: 1246–1247.
52. Fitch WM (1971) Toward Defining the Course of Evolution: Minimum Change for a Specific Tree Topology. *Systematic Zoology* 20: 406–416.
53. Hillis DM, Bull JJ, White ME, Badgett MR, Molineux JJ (1992) Experimental phylogenetics: generation of a known phylogeny. *Science* 255: 589–592.
54. Maddison WP, Maddison DR (1992) *MacClade version 3*. Sunderland: Sinauer Press.
55. Pybus OG, Cochrane A, Holmes EC, Simmonds P (2005) The hepatitis C virus epidemic among injecting drug users. *Infect Genet Evol* 5: 131–139.
56. Enomoto N, Sakuma I, Asahina Y, Kurosaki M, Murakami T, et al. (1996) Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. *N Engl J Med* 334: 77–81.
57. Aurora R, Donlin MJ, Cannon NA, Tavis JE (2009) Genome-wide hepatitis C virus amino acid covariance networks can predict response to antiviral therapy in humans. *J Clin Invest* 119: 225–236.
58. Antaki N, Hermes A, Hadad M, Ftayeh M, Antaki F, et al. (2008) Efficacy of interferon plus ribavirin in the treatment of hepatitis C virus genotype 5. *J Viral Hepat* 15: 383–386.
59. Taylor DR, Shi ST, Romano PR, Barber GN, Lai MM (1999) Inhibition of the interferon-inducible protein kinase PKR by HCV E2 protein. *Science* 285: 107–110.
60. Gale MJ Jr, Korth MJ, Tang NM, Tan SL, Hopkins DA, et al. (1997) Evidence that hepatitis C virus resistance to interferon is mediated through repression of the PKR protein kinase by the nonstructural 5A protein. *Virology* 230: 217–227.
61. Hill AB (1965) The Environment and Disease: Association or Causation? *Proc R Soc Med* 58: 295–300.
62. Han JQ, Barton DJ (2002) Activation and evasion of the antiviral 2'-5' oligoadenylate synthetase/ribonuclease L pathway by hepatitis C virus mRNA. *Rna* 8: 512–525.
63. Han JQ, Wroblewski G, Xu Z, Silverman RH, Barton DJ (2004) Sensitivity of hepatitis C virus RNA to the antiviral enzyme ribonuclease L is determined by a subset of efficient cleavage sites. *J Interferon Cytokine Res* 24: 664–676.
64. Kamal SM, Fehr J, Roesler B, Peters T, Rasenack JW (2002) Peginterferon alone or with ribavirin enhances HCV-specific CD4 T-helper 1 responses in patients with chronic hepatitis C. *Gastroenterology* 123: 1070–1083.
65. Simmonds P (2004) Genetic diversity and evolution of hepatitis C virus—15 years on. *J Gen Virol* 85: 3173–3188.
66. Farci P, Shimoda A, Coiana A, Diaz G, Peddis G, et al. (2000) The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science* 288: 339–344.
67. Cox AL, Mosbrugger T, Mao Q, Liu Z, Wang XH, et al. (2005) Cellular immune selection with hepatitis C virus persistence in humans. *J Exp Med* 201: 1741–1752.
68. Kuntzen T, Timm J, Berical A, Lewis-Ximenez LL, Jones A, et al. (2007) Viral sequence evolution in acute hepatitis C virus infection. *J Virol* 81: 11658–11668.
69. Ray SC, Fanning L, Wang XH, Netski DM, Kenny-Walsh E, et al. (2005) Divergent and convergent evolution after a common-source outbreak of hepatitis C virus. *J Exp Med* 201: 1753–1759.
70. Timm J, Li B, Daniels MG, Bhattacharya T, Reyor LL, et al. (2007) Human leukocyte antigen-associated sequence polymorphisms in hepatitis C virus reveal reproducible immune responses and constraints on viral evolution. *Hepatology* 46: 339–349.
71. Feld JJ, Nanda S, Huang Y, Chen W, Cam M, et al. (2007) Hepatic gene expression during treatment with peginterferon and ribavirin: Identifying molecular pathways for treatment response. *Hepatology* 46: 1548–1563.
72. Sarasin-Filipowicz M, Oakeley EJ, Duong FH, Christen V, Terracciano L, et al. (2008) Interferon signaling and treatment outcome in chronic hepatitis C. *Proc Natl Acad Sci U S A* 105: 7034–7039.
73. Noguchi T, Satoh S, Noshi T, Hatada E, Fukuda R, et al. (2001) Effects of mutation in hepatitis C virus nonstructural protein 5A on interferon resistance mediated by inhibition of PKR kinase activity in mammalian cells. *Microbiol Immunol* 45: 829–840.
74. Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, et al. (2004) Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 303: 327–332.
75. Smith DB, McAllister J, Casino C, Simmonds P (1997) Virus 'quasispecies': making a mountain out of a molehill? *J Gen Virol* 78 (Pt 7): 1511–1519.
76. Dimitrova M, Imbert I, Kieny MP, Schuster C (2003) Protein-protein interactions between hepatitis C virus nonstructural proteins. *J Virol* 77: 5401–5414.
77. Simmonds P, Tuplin A, Evans DJ (2004) Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: Implications for virus evolution and host persistence. *Rna* 10: 1337–1351.
78. Ogata N, Alter HJ, Miller RH, Purcell RH (1991) Nucleotide sequence and mutation rate of the H strain of hepatitis C virus. *Proc Natl Acad Sci U S A* 88: 3392–3396.
79. Kato N (2000) Genome of human hepatitis C virus (HCV): gene organization, sequence diversity, and variation. *Microb Comp Genomics* 5: 129–151.
80. Lloyd AR, Jagger E, Post JJ, Crooks LA, Rawlinson WD, et al. (2007) Host and viral factors in the immunopathogenesis of primary hepatitis C virus infection. *Immunol Cell Biol* 85: 24–32.
81. Farci P, Shimoda A, Wong D, Cabezon T, De Gioannis D, et al. (1996) Prevention of hepatitis C virus infection in chimpanzees by hyperimmune serum against the hypervariable region 1 of the envelope 2 protein. *Proc Natl Acad Sci U S A* 93: 15394–15399.
82. Munoz de Rueda P, Casado J, Paton R, Quintero D, Palacios A, et al. (2008) Mutations in E2-PePHD, NS5A-PKRBD, NS5A-ISDR, and NS5A-V3 of hepatitis C virus genotype 1 and their relationships to pegylated interferon-ribavirin treatment responses. *J Virol* 82: 6644–6653.
83. Berg T, Mas Marques A, Hohne M, Wiedenmann B, Hopf U, et al. (2000) Mutations in the E2-PePHD and NS5A region of hepatitis C virus type 1 and the dynamics of hepatitis C viremia decline during interferon alfa treatment. *Hepatology* 32: 1386–1395.
84. Gaudy C, Lambele M, Moreau A, Veillon P, Lunel F, et al. (2005) Mutations within the hepatitis C virus genotype 1b E2-PePHD domain do not correlate with treatment outcome. *J Clin Microbiol* 43: 750–754.
85. Shiffman ML (2006) Chronic hepatitis C: treatment of pegylated interferon/ribavirin nonresponders. *Curr Gastroenterol Rep* 8: 46–52.
86. Shiffman ML (2008) Optimizing the current therapy for chronic hepatitis C virus: peginterferon and ribavirin dosing and the utility of growth factors. *Clin Liver Dis* 12: 487–505, vii.
87. Dal Pero F, Tang KH, Gerotto M, Bortoletto G, Paulon E, et al. (2007) Impact of NS5A sequences of Hepatitis C virus genotype 1a on early viral kinetics during treatment with peginterferon- alpha 2a plus ribavirin. *J Infect Dis* 196: 998–1005.
88. Cannon NA, Donlin MJ, Fan X, Aurora R, Tavis JE (2008) Hepatitis C virus diversity and evolution in the full open-reading frame during antiviral therapy. *PLoS ONE* 3: e2123.
89. von Wagner M, Lee JH, Ruster B, Kronenberger B, Sarrazin C, et al. (2003) Dynamics of hepatitis C virus quasispecies turnover during interferon-alpha treatment. *J Viral Hepat* 10: 413–422.
90. Rubbia-Brandt L, Fabris P, Paganin S, Leandro G, Male PJ, et al. (2004) Steatosis affects chronic hepatitis C progression in a genotype specific way. *Gut* 53: 406–412.
91. Rumi MG, De Filippi F, La Vecchia C, Donato MF, Gallus S, et al. (2005) Hepatitis C reactivation in patients with chronic infection with genotypes 1b and 2c: a retrospective cohort study of 206 untreated patients. *Gut* 54: 402–406.
92. Coppola N, Vatiere LM, Sagnelli E (2005) HCV genotype 2 as a risk factor for reactivation of chronic HCV infection. *Gut* 54: 1207.

93. Gao B, Hong F, Radaeva S (2004) Host factors and failure of interferon-alpha treatment in hepatitis C virus. *Hepatology* 39: 880–890.
94. Khakoo SI, Thio CL, Martin MP, Brooks CR, Gao X, et al. (2004) HLA and NK cell inhibitory receptor genes in resolving hepatitis C virus infection. *Science* 305: 872–874.
95. He XS, Ji X, Hale MB, Cheung R, Ahmed A, et al. (2006) Global transcriptional response to interferon is a determinant of HCV treatment outcome and is modified by race. *Hepatology* 44: 352–359.
96. Holmes EC (2008) Evolutionary history and phylogeography of human viruses. *Annu Rev Microbiol* 62: 307–328.
97. Kuiken C, Hraber P, Thurmond J, Yusim K (2008) The hepatitis C sequence database in Los Alamos. *Nucleic Acids Res* 36: D512–516.
98. McOmish F, Yap PL, Dow BC, Follett EA, Seed C, et al. (1994) Geographical distribution of hepatitis C virus genotypes in blood donors: an international collaborative survey. *J Clin Microbiol* 32: 884–892.
99. Jaeckel E, Cornberg M, Wedemeyer H, Santantonio T, Mayer J, et al. (2001) Treatment of acute hepatitis C with interferon alfa-2b. *N Engl J Med* 345: 1452–1457.
100. Zekry A, Patel K, McHutchison JG (2005) Treatment of acute hepatitis C infection: more pieces of the puzzle? *J Hepatol* 42: 293–296.
101. Fishman SL, Murray JM, Eng FJ, Walewski JL, Morgello S, et al. (2008) Molecular and bioinformatic evidence of hepatitis C virus evolution in brain. *J Infect Dis* 197: 597–607.
102. Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, et al. (2002) Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med* 347: 975–982.
103. Dalgard O, Bjoro K, Ring-Larsen H, Bjornsson E, Holberg-Petersen M, et al. (2008) Pegylated interferon alfa and ribavirin for 14 versus 24 weeks in patients with hepatitis C virus genotype 2 or 3 and rapid virological response. *Hepatology* 47: 35–42.
104. von Wagner M, Huber M, Berg T, Hinrichsen H, Rasenack J, et al. (2005) Peginterferon-alpha-2a (40KD) and ribavirin for 16 or 24 weeks in patients with genotype 2 or 3 chronic hepatitis C. *Gastroenterology* 129: 522–527.
105. Derbala M, Amer A, Bener A, Lopez AC, Omar M, et al. (2005) Pegylated interferon-alpha 2b-ribavirin combination in Egyptian patients with genotype 4 chronic hepatitis. *J Viral Hepat* 12: 380–385.
106. Kamal SM, El Tawil AA, Nakano T, He Q, Rasenack J, et al. (2005) Peginterferon {alpha}-2b and ribavirin therapy in chronic hepatitis C genotype 4: impact of treatment duration and viral kinetics on sustained virological response. *Gut* 54: 858–866.
107. El-Zayadi AR, Attia M, Barakat EM, Badran HM, Hamdy H, et al. (2005) Response of hepatitis C genotype-4 naive patients to 24 weeks of Peg-interferon-alpha2b/ribavirin or induction-dose interferon-alpha2b/ribavirin/ amantadine: a non-randomized controlled study. *Am J Gastroenterol* 100: 2447–2452.
108. Alfaleh FZ, Hadad Q, Khuroo MS, Aljumah A, Algamedy A, et al. (2004) Peginterferon alpha-2b plus ribavirin compared with interferon alpha-2b plus ribavirin for initial treatment of chronic hepatitis C in Saudi patients commonly infected with genotype 4. *Liver Int* 24: 568–574.
109. Hasan F, Asker H, Al-Khaldi J, Siddique I, Al-Ajmi M, et al. (2004) Peginterferon alfa-2b plus ribavirin for the treatment of chronic hepatitis C genotype 4. *Am J Gastroenterol* 99: 1733–1737.